Crowdsourcing Mental Models using DESIM (Descriptive to Executable Simulation Modeling)

Mark S. PFAFF, Jill L. DRURY, and Gary L. KLEIN
The MITRE Corporation

ABSTRACT

This paper describes the DESIM (Descriptive to Executable Simulation Modeling) process for transforming causal descriptive models into computer simulation models based on information obtained from crowdsourcing. Feedback obtained from crowdsourcing is used to quantify the strength of causal relationships between variables in descriptive models to provide consistent and less-biased distributions of estimated weights for each causal relationship and thereby enable mathematical processing of the descriptive models on a computer. The approach employs fuzzy cognitive modeling methods to elicit and structure the models and the analytic hierarchy process to compute the distribution of weights between variables. The output of this process produces a decision space, which is visualized with a novel decision space visualization tool. An experimental application of this process is presented and discussed, with implications for future research.

KEYWORDS

Mental models; General and miscellaneous; Judgment and Decision Making; Planning and Prediction; Expertise

INTRODUCTION

There is increasing research on better ways to support decision makers when they need to choose among options in complex situations. Because of the deep uncertainty surrounding such situations (Walker, Lempert, & Kwakkel, 2012), questions arise about planning for the range of conditions under which reasonable operations would be possible, or determining what operations may be managed under current conditions (Caldwell, 2014). In some situations, decision makers match salient cues presented by the external environment to a mental template built from previous experiences: part of their mental model (Craik, 1943). They then envision at least one possible course of action and mentally simulate the results of applying that action to determine whether that option is acceptable (Klein, 1998). This process is more difficult to use in complex situations, or by novice decision makers who have not yet acquired sufficient experience to map the current situation to mental templates showing successful resolutions of problems faced in the past. Addressing this gap are *decision spaces*, defined as the range of options, the underlying interconnected factors that influence their relative desirability, and the landscape of plausible futures that could accompany any given course of action (Pfaff et al., 2012). For example, alternative courses of action can be evaluated using a computer simulation, which can provide decision makers with a graphical depiction of their decision space and thereby option awareness during real-world emergency situations, such as a natural disaster. The idea is that visually depicting decision spaces can be like providing night vision goggles for the mind, offloading the mental simulation process onto the computer, which displays the results of many posssible options using an intuitive decision space visualization (DSV) that otherwise cannot be seen unaided.

In laboratory experiments, DSVs have enabled decision makers to make choices faster, more accurately, and with more confidence than without the DSV (Pfaff et al., 2012). The process of creating DSVs relies upon exploratory modeling (Bankes, 1993; Chandrasekaran & Goldman, 2007; Chandrasekaran, 2005). In exploratory modeling, analysts construct a set of plausible assumptions about the environment in which a decision will be made, run a simulation model that includes a parameterization of those assumptions, and score the outcomes for each decision option according to one or more evaluative criteria. The analyst varies each of the parameters representing the assumptions to account for uncertainty, and runs the model repeatedly to obtain a range of outcomes for each decision option. DSVs consist of a frequency-based depiction of the range of outcomes for each option. Thus, the process of constructing a DSV requires executing a model that pertains to the domain and situation encountered by the decision maker. While there is a rich history of research into modeling and simulation, the fact remains that developing validated models can be costly, time-consuming, and error-prone. The need for models has become the stumbling block for creating DSVs for broad classes of decision making situations.

There is a need for a more streamlined way to develop models. Building from decision makers' mental simulation abilities, new research in crowdsourcing points to the promise of combining the mental models from multiple decision makers to paint a more complete and less biased picture of a situation than any individual might be able to achieve in isolation. The result is a new process, DESIM (Descriptive to Executable Simulation Modeling), that can create a computational model of a situation in hours or days instead of weeks or months.

DESIM consists of the following stages:
- Create one or more validated descriptive causal models
- Deconstruct the model into pairwise comparisons
- Crowdsource the comparisons and compute relationship strengths
- Apply the computational model to create DSVs

The DESIM process transforms descriptive causal models into computer simulation models based on information obtained from crowdsourcing. A computer user interface for crowdsourcing, combined with computational algorithms, produces quantitative values for the strengths of causal relationships between variables in the descriptive models, resulting in less biased distributions of estimated values for each relationship, and enabling the models to be computationally processed. This system generates improved outcome spaces, which refer to one or more possibilities regarding the relationships among options, actions, or variables that can be used to analyze the subject of a computer model. For example, a decision space can be an outcome space used by decision makers to determine how to respond to a complex situation based on the relationships between options and their plausible effects that can be forecasted by a computer simulation from facts about the situation. A key distinguishing feature of this system from other decision support tools is that it presents to the user an interactive and dynamic frequency distribution of possible outcomes (e.g. box-plot or histogram) rather than a single static probability, which conceals important knowledge about the range and distribution of possible outcomes. For example, perceiving a distribution with a long tail or a bi-modal distribution may lead to a significantly different decision (or to further exploration of the data) than simply knowing the mean probability of success. Moreover, in this format, further exploration of the data can yield deeper awareness of what factors lead to better vs. worse outcomes (Drury, Klein, Musman, Liu, & Pfaff, 2012). The rest of this paper describes how the DESIM process works, prefaced by related work and ending with a brief example of using DESIM.

RELATED WORK

Conventional computer simulation systems include models that are designed based on expert knowledge for use to simulate different situations that may occur in the real-world. These computer simulations are assumed to be reliable because they are created using expert knowledge. Unfortunately, each expert has certain behavioral patterns, preferences, and characteristics that may bias the programming of models. For example, different experts may agree to include certain variables in a particular computer model but disagree about the significance of each variable. Thus, conventional computer simulations created using expert models may be biased and unreliable, and the process of translating expert knowledge into computer simulation models can be slow and error prone (Bankes, 1993). Because this approach most often requires computer programmers to do this translation, the process is slow and tedious since the translation must be carefully and constantly validated by the experts to eliminate translation errors.

Alternatively, a domain expert's (such as an analyst or forecaster) descriptive causal model can be elicited for a focal question such as "Will Iran invade the Strait of Hormuz?" and then represented in a digital data structure that is interpretable by computer programs and, moreover, can be displayed in a graphical presentation that is easy for the expert to validate. This is the approach taken with DESIM. Most often, multiple domain experts are interviewed to understand different and potentially conflicting perspectives on the problem. Experts identify model components, links between components, and the dynamic and functional relationships among the components. The advantages of this process include the ability to capture perceptions that are difficult to quantify and the participatory form makes data collection more approachable and engaging for domain experts not familiar with modeling processes (Özesmi & Özesmi, 2004). An alternative approach is to separate the work of developing the formal model from the interview process, such that the model itself is designed by modeling experts, based on the knowledge collected from interviews with domain experts. The resulting model is then validated by displaying a graphical representation to the domain experts who check it for completeness and accuracy (Sieck, Rasmussen, & Smart, 2010).

There are multiple tools for computationally representing mental models. The model in Figure 1, below, was designed using CMapTools (Cañas et al., 2004), which provides a graphical interface for constructing and editing cognitive models and provides machine-readable output for use by other computational tools. The edge labels are an open text field which here is used to signify positive or negative associations between nodes. A similar product is MentalModeler (Gray, Gray, Cox, & Henly-Shepard, 2013), designed to support the fuzzy cognitive modeling process from model elicitation and graphical representation, variable edge weight selection, to exploratory simulation (it also can export the model in a machine-readable form).

DESIM also uses crowdsourcing (or crowd estimating) to quantify the relationships in the descriptive model Crowdsourcing is a process of obtaining services, ideas, or content by soliciting contributions, especially over the Internet, from a large group of people referred to as a crowd (Howe, 2006). This process typically involves a division of labor for tedious tasks split among members of the crowd. For example, crowdsourcing can be used to solicit predictions for a political campaign, or to search for answers, solutions, or a missing person (Surowiecki, 2005). In other words, crowdsourcing combines the incremental efforts of numerous contributors to achieve a greater result in a relatively short period of time. Lin et al. (2012) used crowdsourcing to understand mental models of privacy in mobile applications, but did not create a model explicitly.

METHODS

This section details the methods for eliciting the canonical cognitive model of the problem, representing the model interactively, eliciting values for the model from the crowd, and analyzing and visualizing the resulting data. In the first stage, analysts develop a focal question of interest and interview one or a few experts on the subject area to elicit the experts' mental models of the factors they believe influence the outcomes of the focus question. The analysts develop one causal model per expert in the form of nodes and unweighted edges, validating each model with its expert. The causal mental models may be similar or may diverge. Analysts combine multiple models when they are mathematically equivalent, but it is acceptable to have more than one canonical model describing the experts' mental models.

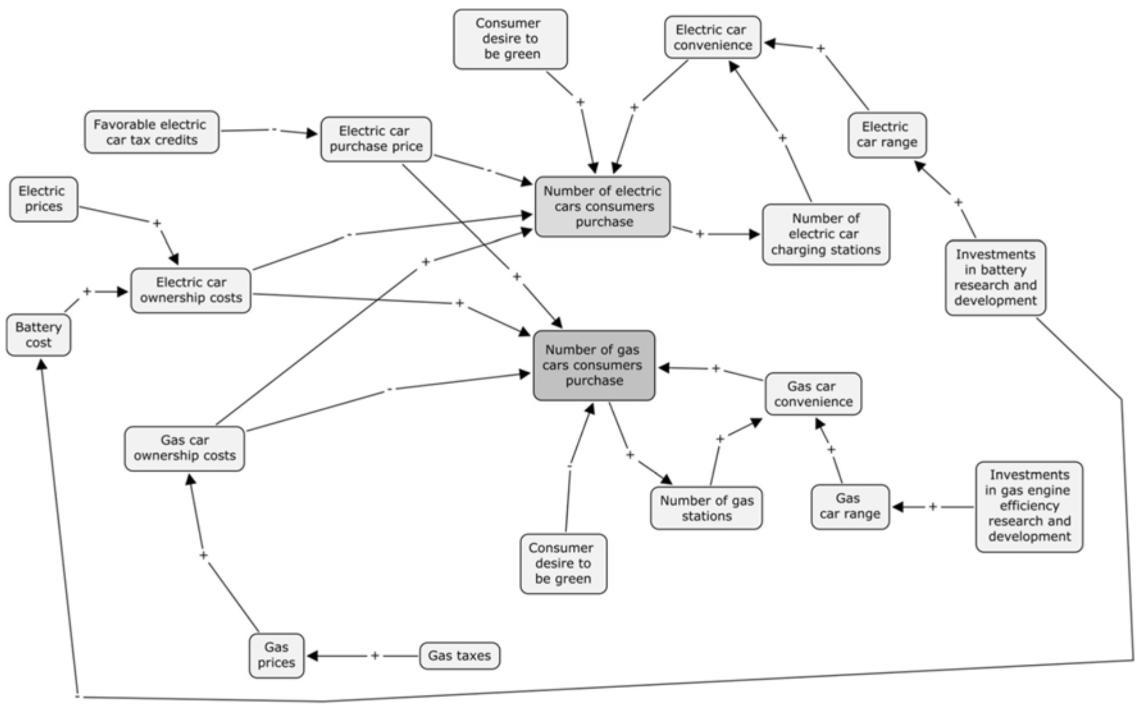

Figure 1. Example cognitive model describing reasons to purchase a gas or electric vehicle

The descriptive causal model can be represented on a computer as a graph of nodes connected by edges, as shown in Figure 1, for which the focal question asked of domain experts was "Will consumers buy more electric vehicles than gas vehicles in 2018?" A node in a descriptive model is a variable that represents a concept such as an action, option, or policy that has a range of values. Different scenarios using the model would be described with different sets of initial node values. An edge includes a weight that represents a causal association or relationship between two or more nodes. The sign of an edge weight denotes a direction of correlation between nodes, and the magnitude of an edge weight denotes the strength of the causal relationship between the nodes. While a static value for an edge weight may be elicted from a single expert, a distribution of values for the edge weights can be determined through appropriately crowdsourcing to multiple experts. The experts may or may not include those who intially described the model, depending on whether the crowd of experts should be agnostic of the full model and be focused solely on comparing pairs of concepts regardless of context. An algorithm can be used to determine how crowd sourced feedback defines the distribution of edge weights.

Once the canonical map is elicited from one or more domain experts and represented structurally, the DESIM computer program processes it in parts in order to obtain edge weights. While experts are able to give the sign (+ or -) of a causal relationship, they are less able to give an accurate estimate of the magnitude (Osei-Bryson, 2004). Because subjective point estimates are unreliable, another method is necessary to produce accurate edge weights. This is achieved through a systematic set of pairwise comparisons of the connected node pairs in the model, where an expert rates the comparative strength of two relationships (e.g. whether relationship $X_1 \rightarrow X_2$ is stronger than $X_3 \rightarrow X_4$, and by how much). In this study, crowdsourcing of the pairwise comparisons was used via Amazon Mechanical Turk (AMT) because the topic was of general interest and required no special expertise. More specialized models would require targeted recruitment of subject-matter experts, for example, from a community of analysts or forecasters in a specific field.

A web-based interface was built to elicit the online pairwise comparisons. This tool takes as input the machine-readable model produced in the preceding steps and generates the set of pairwise comparisons which are presented in sequence to the users recruited via AMT. First, a single relationship $X_1 \rightarrow X_2$ is graphically presented to the user with the question "Do you agree with this relationship?" The three choices in this example are "Agree: An increase on the left causes an increase on the right", "Disagree: An increase on the left has no effect on the right", or "Disagree: An increase on the left causes a decrease on the right" (see Figure 2).

After the user has agreed with at least two relationships in the model, two relationships ($A = X_1 \rightarrow X_2$ and $B = X_3 \rightarrow X_1$) are presented with the question "Which relationship is stronger?" with the choices "A is stronger than B," "B is stronger than A," or "A is the same as B." (see Figure 3). If either of the first two choices are selected, the user is additionally asked "How much stronger?" and presented with a slider ranging from "A is much stronger than B" to "A is the same as B." After answering, the user then proceeds to the next comparison. When the user disagrees with a given relationship, it is given a weight of zero and eliminated from all future pairwise comparisons. Because different users may agree or disagree with different relationships, the resulting set of pairwise comparisons will vary in their degree of completeness.

DESIM analyzes the results of the pairwise comparisons using the Analytic Hierarchy Process (AHP; Saaty, 1990). From each respondent's set of pairwise comparisons, a set of edge weights is computed using a modified

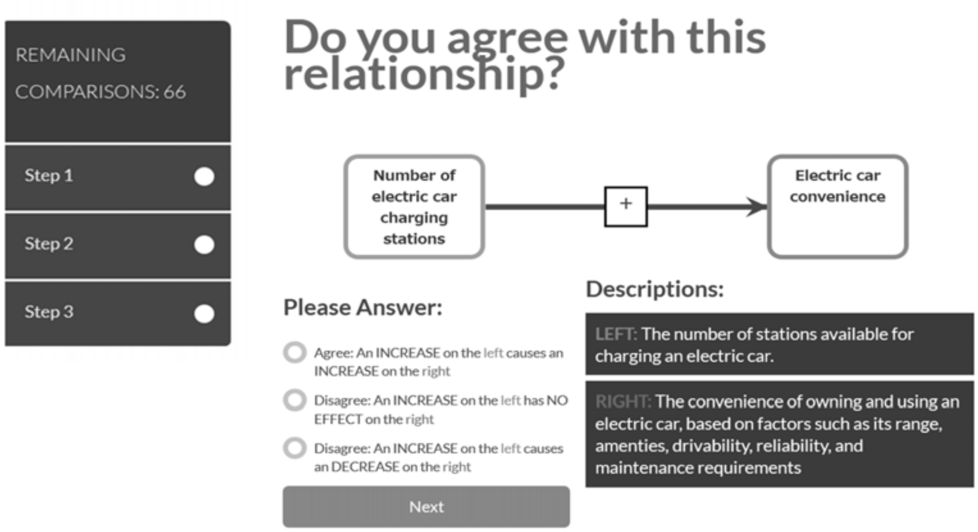

Figure 2. Step 1 of causal relationship pairwise comparison

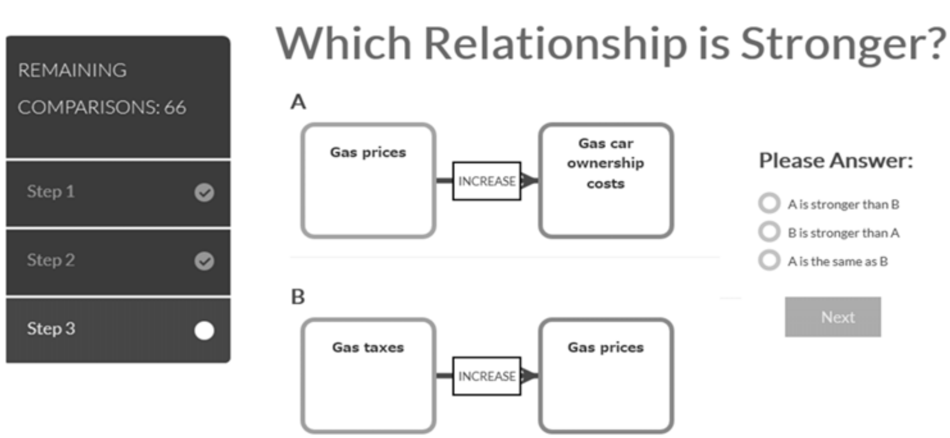

Figure 3. Step 2 of causal relationship pairwise comparison

AHP technique to accommodate incomplete sets of pairwise comparisons (Harker, 1987). The sets of edge weights for all respondents (which are values between 0 and 1; the model defines the sign) are aggregated and used to populate the original model with distributions of edge weights for each relationship in the model.

Using these distributions of weights, multiple simulation model processing runs can be performed to generate one or more outcome spaces, using an iterative Fuzzy Cognitive Modelling (FCM) method (Kosko, 1986). Initial node values and edge weights can be varied for each model processing run to create an outcome node distribution. Variations may be generated in different ways. For example, in the example presented here, the model was executed once for each set of edge weights elicted from the respondents. Alternately, a Monte-Carlo method can be used to generate each variation by sampling from distributions of node values and edge weights. An analysis of the resulting outcome space provides a more comprehensive understanding than a single aggregated mean estimate about how various variables impact consumer interest in electric cars.

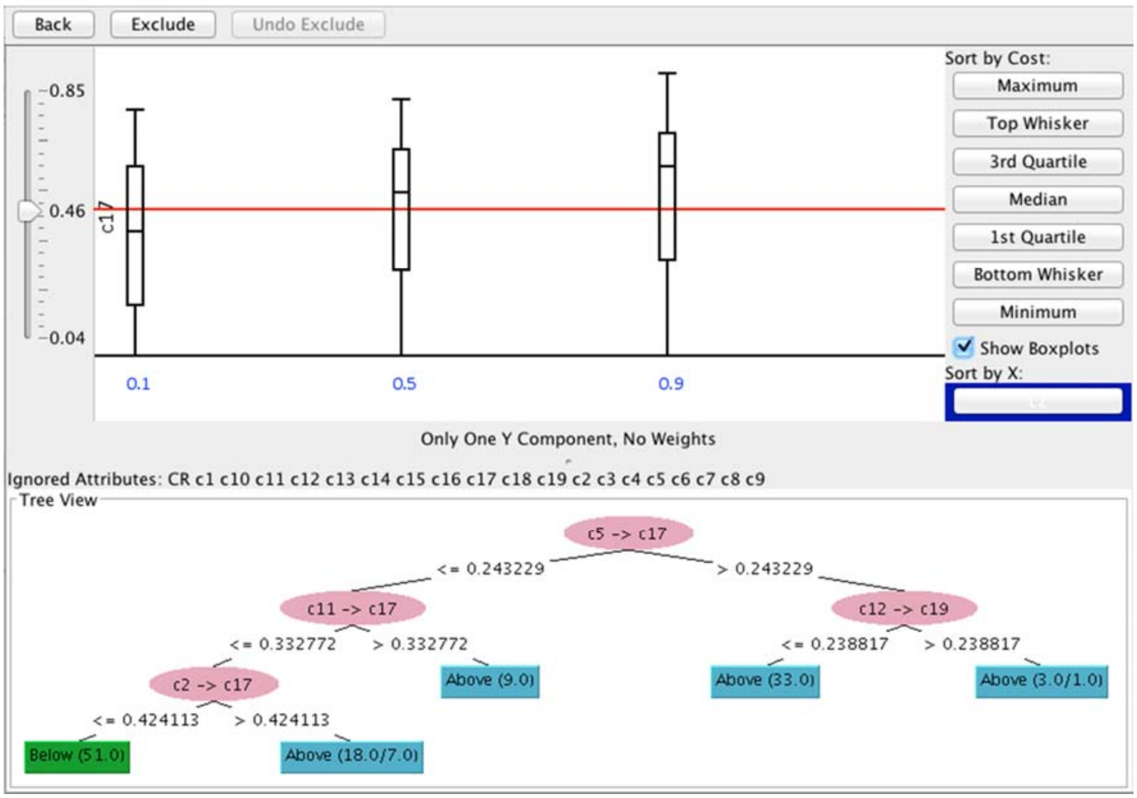

Figure 4. Decision space visualization of the effect of low, medium, and high consumer desire to be green affects the number of electric powered cars consumers purchase

The outcome space can then be represented by a decision space visualization (DSV; Figure 4). In the top portion, the X-axis corresponds to each permutation of a scenario or course-of-action option, and the Y-axis corresponds to the value of an outcome node. For multiple outcome nodes, the Y-axis may include multiple measures for a weighted composite value. This DSV displays the distribution of outcomes for each potential decision option under various plausible conditions. For example, in the case where the Y-axis is cost, and higher cost is bad, options with a low and tight distribution represent more robust options, those which are likely to turn out well under a wide variety of conditions. Options with broader distributions show higher sensitivity to conditions in the model and warrant caution. The tree diagram on the bottom portion of the display, generated using the WEKA package (Hall et al., 2009), represents a hierarchy of underlying interacting characteristics explaining the outcomes above or below the threshold selected in the top portion (the red line at 0.46). The tree is ordered in descending level of influence on the outcomes under inspection. This interactive visualization helps the decision maker understand which nodes in the model are more or less influential on given outcomes in the decision space. By manipulating the horizontal line on the upper diagram to change the threshold of what is considered to be desirable versus undesirable outcomes, this visualization approach allows a decision maker to actually see relationships between options that are otherwise obscured, rather than mentally simulating each one.

We conducted a study to validate the DESIM system in practice. The car-buying model described above was divided into 22 node pairs and the pairwise comparison process described above was administered to 38 workers recruited from Amazon Mechanical Turk, who were paid $4 for their time. Data was collected in two batches.

The first 11 individuals responded in 90 minutes after releasing the task. Data collection was closed temporarily to verify the integrity of the incoming data, and then reopened two days later for two hours, collecting another 27 responses. AHP analysis (Harker, 1987) calculated the distributions of edge weights from the respondents' pairwise comparisons. The Java FCM library (De Franciscis, 2014) was used to compute outcome values for the proportion of gas and electric cars consumers purchase (numbers between 0 and 1), for three different scenarios (low, medium, and high consumer desire to be green), with all other input nodes held constant. Consumer desire to be green refers to an individual's preference to make decisions intended to benefit the environment. The FCM was computed for each of the 38 sets of edge weights, resulting in distributions of 38 outcomes for each of the three scenarios, which are shown in the top portion of Figure 4.

RESULTS

In Figure 4, the outcome of interest is node C17, the number of electric cars consumers purchase. Preliminary analysis of the outcomes showed them to be bimodal, so the threshold is set to 0.46 to differentiate the top half from the bottom half of the outcomes for the desire to buy electric cars (the boxplots may be toggled off to show the raw data points, not shown here).

It is clear that as the consumer's desire to be green increases, so does the prediction for the median number of electric cars consumers will purchase, as expected. However, the bottom portion of Figure 4 helps explore the bimodality in the data. The factors under consideration are the various edge weights provided by the crowdsourced population described above. The bottom display has calculated the rules explaining what makes outcomes score above or below the threshold of 0.46, across all three values of the desire to be green. According to the tree display, the most important discriminating factor is the edge between nodes C5 (electric car ownership costs) and C17 (number of electric cars consumers purchase). When an expert rates the weight of this edge greater than 0.24, all outcomes are above the threshold. However, when this edge is rated less than 0.24, the next most influential edge is the relationship between C11 (gas car ownership costs) and C17. When this edge is rated greater than 0.33, the outcomes are above the threshold. Finally, if that edge is rated less than 0.33, it comes down to the expert's rating of the edge between C2 (consumer desire to be green) and C17. Therefore, what explains the outcomes below the threshold are that they are the opinions of experts who believe that all three edges mentioned above have weights below the three indicated tipping points. It also indicates in descending order which of the relationships are most influential and therefore the ones most worthy of attention for decision making. Not shown in this diagram is the ability to select one or more specific outcomes in the top portion of the display, which highlights the corresponding leaves in the tree in the bottom half; the reverse is also possible (Drury, et al., 2012).

CONCLUSION

In summary, the DESIM system elicits mental causal descriptive models from people and transforms them into computer processible causal simulation models, which allows for offloading the simulation-modeling burden from people to the computer. Consequently, by returning choice to a perceptual comprehension process in a decision space visualization, this approach enables decision makers to apply their more powerful visual pattern matching capabilities, rather than their more limited capacities for mental simulation.

Our future work will apply a web-based crowdsourcing system to participatory descriptive model development so that interviews and manual model creation will no longer be necessary. A similar existing system is Scheherezade (Li, Lee-Urban, & Riedl, 2012), a crowdsourcing tool that elicits domain knowledge to create causal narrative models on a given topic, called "plot-graphs." It relies on a structured natural-language processing approach to develop a narrative diagram, represented as a directed acyclic graph. Turkomatic (Kulkarni, Can, & Hartmann, 2012) uses a related aproach to have crowd workers break down a given task into a detailed workflow, making it suited for representing procedural knowledge in the form of a causal model. The contribution of our work, as the first project to use crowdsourced mental (causal descriptive) models that translate into computer-based simulation models, is a streamlined route to model-based decision support tools.

ACKNOWLEDGMENTS

We thank Arthur Liu and Scott Musman for development of the decision space visualizer tool, and Dimitri De Franciscis for support with the Java FCM library. This work was partially supported by the MITRE Corporation Innovation Projects 19MSR080-BA under the Systems Engineering Investment Area and 19MSR083-CA under the Data to Decisions Investment Area.

REFERENCES

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. doi:10.1287/opre.41.3.435

Caldwell, B. S. (2014). Cognitive challenges to resilience dynamics in managing large-scale event response. *Journal of Cognitive Engineering and Decision Making*, *8*(4), 318–329. doi:10.1177/1555343414546220

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., & Gómez, G. (2004). CmapTools: A knowledge modeling and sharing environment. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping* (pp. 125–133). Pamplona, Spain: Universidad Pública de Navarra.

Chandrasekaran, B. (2005). *From optimal to robust COAs: Challenges in providing integrated decision support for simulation-based COA planning*. Columbus, OH: Ohio State University, Laboratory for AI Research.

Chandrasekaran, B., & Goldman, M. (2007). Exploring robustness of plans for simulation-based course of action planning: A framework and an example. In *Paper presented at the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making* (pp. 185–192). Honolulu, HI.

Craik, K. J. W. (1943). *The nature of explanation.* Oxford: University Press, Macmillan.

De Franciscis, D. (2014). JFCM: A Java library for fuzzy cognitive maps. In E. I. Papageorgiou (Ed.), *Fuzzy cognitive maps for applied sciences and engineering* (Springer-V., pp. 199–220). Berlin Heidelberg.

Drury, J. L., Klein, G. L., Musman, S., Liu, Y., & Pfaff, M. (2012). Requirements for data mining the decision space. In *Proceedings of the 2012 International Command and Control Research and Technology Symposium*. Fairfax, VA.

Gray, S., Gray, S., Cox, L. J., & Henly-Shepard, S. (2013). Mental Modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. *2013 46th Hawaii International Conference on System Sciences*, 965–973. doi:10.1109/HICSS.2013.399

Hall, M., Frank, E., Holmes, G., Pfaringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1).

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, *9*(11), 837–848. doi:10.1016/0270-0255(87)90503-3

Howe, J. (2006, June). The rise of crowdsourcing. Retrieved December 12, 2014, from http://archive.wired.com/wired/archive/14.06/crowds_pr.html

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, *24*(1), 65–75.

Kulkarni, A., Can, M., & Hartmann, B. (2012). Collaboratively crowdsourcing workflows with Turkomatic. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12* (pp. 1003–1012). New York, New York, USA: ACM Press. doi:10.1145/2145204.2145354

Li, B., Lee-Urban, S., & Riedl, M. O. (2012). Toward autonomous crowd-powered creation of interactive narratives. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference* (pp. 20–25). Stanford University.

Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J. I., & Zhang, J. (2012). Expectation and purpose. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12* (pp. 501–510). New York: ACM Press. doi:10.1145/2370216.2370290

Osei-Bryson, K.-M. (2004). Generating consistent subjective estimates of the magnitudes of causal relationships in fuzzy cognitive maps. *Computers & Operations Research*, *31*(8), 1165–1175. doi:10.1016/S0305-0548(03)00070-4

Özesmi, U., & Özesmi, S. L. (2004). Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, *176*(1-2), 43–64. doi:10.1016/j.ecolmodel.2003.10.027

Pfaff, M. S., Klein, G. L., Drury, J. L., Moon, S. P., Liu, Y., & Entezari, S. O. (2012). Supporting complex decision making through option awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 155–178. doi:10.1177/1555343412455799

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*(1), 9–26. doi:10.1016/0377-2217(90)90057-I

Sieck, W. R., Rasmussen, L. J., & Smart, P. R. (2010). Cultural network analysis: A cognitive approach to cultural modeling. In D. Verma (Ed.), *Network Science for Military Coalition Operations: Information Extraction and Interaction* (pp. 237–255). Hershey, PA: IGI Global.

Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Anchor Books.

Walker, W. E., Lempert, R. J., & Kwakkel, J. H. (2012). *Deep uncertainty*. Retrieved from http://www.hau.gr/resources/toolip/doc/2012/05/10/deep-uncertainty_warren-e-walker.pdf